Separation Properties and Related Bounds of Collusion-secure Fingerprinting Codes

Myong-Son SIHN and Ryul KIM

Faculty of Mathematics, Kim Il Sung University, Pyongyang, D. P. R. Korea
Email address : ryul_kim@yahoo.com

Abstract

In this paper, we investigate separation properties and related bounds of some codes. Firstly, we obtain an existence result for $(w_1, w_2)-$separating codes and discuss the "optimality" of the upper bounds. Secondly, we propose an upper bound on the size of restricted $(w, w)-$separating codes. Thirdly, we integrate former results and some new conditions for Silverberg's open problem into an algebraic version, and end with an interesting relationship between separation and existence of non-trivial subspace subcode of Reed-Solomon codes.

Keywords: Fingerprinting, Silverberg's Problem, Subspace Subcode

1 Introduction

1.1 Digital Fingerprinting

The growth of Internet raised the problem of illegal redistribution as a major concern in digital content industry, because copying such material is easy and no information is lost in the process. To protect digital copies, however, is a complicated task. Methods like cryptography do not resolve this problem, since the information must be decrypted at one point to be able to use it. The goal of digital fingerprinting is to discourage people from illegally redistributing their legally purchased copy. In this scenario, the distributor embeds into the digital content, using a watermark algorithm, a unique piece of information (*fingerprint*) for each user. If an illegal copy is found, the distributor can extract the fingerprint from it to identity a dishonest user (*pirate*). Because the pirate may try to damage the fingerprint before redistribution, the watermarking algorithm must ensure robustness to the distributor.

Nevertheless, the most dangerous attack against digital fingerprinting is a *collusion attack*. The contents delivered to different users are, since their fingerprints differ, essentially different. Two or more pirates may compare their copies and reveal the locations of part of fingerprint. With deleting or modifying those locations, pirates can generate a new copy of content in order not to be traced. This collusion attack could not only violate pirate-identifying but frame an innocent user in some cases. We are interested in

designing a set of fingerprints(*fingerprinting code*) with which the distributor can always identity at least one colluder from a forged fingerprint with a small error probability.

1.2 Definitions and Basic Results

We will denote the ith component of any tuple x by x_i and the Hamming distance between two tuples x, y by $d(x, y)$.

A set $\Gamma \subset \mathbb{F}_q^n$ is called a $q-$*ary* $(n, M)-$*code* or $(n, M)_q-$*code* if $|\Gamma| = M$. An $[n, k]_q-$*linear code* is a linear subspace of $\mathbb{F}_q^n$. $R := n^{-1}\log_q M$ is called a *rate* of code. For an arbitrary nonempty subset U of a $q-$ary code, we define *detectable position set*, *descendant set* and *feasible set* by the followings.

$$\begin{aligned}
&\text{det}U := \{i \in \overline{1, n} \mid \exists a, b \in U : a_i \neq b_i\} \\
&\text{desc}U := \{x \in \mathbb{F}_q^n \mid \forall i, \exists a \in U : x_i = a_i\} \\
&\text{fea}U := \{x \in \mathbb{F}_q^n \mid i \notin \text{det}U \Rightarrow x_i = a_i, a \in U\}
\end{aligned}$$

We now give the following definitions concerning collusion-secure properties of codes.

Definition 1.1 Suppose Γ is an $(n, M)_q-$code and w, w_1, w_2 are positive integers.

- Γ is a $(w_1, w_2)-$*separating code* provided that, if U_1, U_2 are disjoint subsets of Γ such that $1 \leq |U_1| \leq w_1$ and $1 \leq |U_2| \leq w_2$, then their descendant sets are also disjoint.
- Γ is a *restricted* $(w_1, w_2)-$*separating code* provided that, if U_1, U_2 are disjoint subsets of Γ such that $1 \leq |U_1| \leq w_1$ and $1 \leq |U_2| \leq w_2$, then their feasible sets are also disjoint.
- $w-$*frameproof code*(*FP code*, shortly) is a $(w, w)-$ separating code.
- Γ is called a $w-$*identifiable parent property code*(*IPP code*, shortly) provided that for all $x \in \mathbb{F}_q^n$, the set $\text{IPP}_w(x) := \{U \subset \Gamma \mid x \in \text{desc}U, 1 \leq |U| \leq w\}$ is empty or $\bigcap_{U \subset \text{IPP}_w(x)} U \neq \phi$.
- Γ is called a $w-$*traceability code*(*TA code*, shortly) provided that, if $U \subset \Gamma, 1 \leq |U| \leq w$ and $x \in \text{desc}U$, there is at least one codeword $y \in U$ such that $d(x, y) < d(x, z)$ for all $z \in \Gamma \backslash U$.

The intuitive meanings of FP, SFP, IPP and TA properties are as follows.

- A code is $w-$FP if no coalition of size at most w can frame another user not in the coalition by forging the codeword of that user.
- A code is $w-$SFP if no coalition of size at most w can frame a disjoint coalition of size at most w by forging a codeword which could have been produced by the second coalition.

- A code is $w-$IPP if no coalition of size at most w can forge a codeword that cannot be traced back to at least one member of the coalition.
- A $w-$TA code is a $w-$IPP code with a fast tracing algorithm based on Hamming distance.

The code classes stated above are known to satisfy the following relationships.

Proposition 1.1 (see [12]) *Let d be the minimum distance of $(n, M)_q-$code Γ and $w \geq 2$ be a positive integer. Then for Γ,*

$$d > n(1 - 1/w^2) \Rightarrow w - TA \Rightarrow w - IPP \Rightarrow w - SFP \Rightarrow w - FP$$

Proposition 1.2 (see [2]) *Let d be the minimum distance of $(n, M)_q-$code Γ and w_1, w_2 be positive integer. If $d > n(1 - 1/(w_1 w_2))$, then Γ is a $(w_1, w_2)-$ separating code.*

Separating codes have been studied with combinatorial methods in many works. An $(N; n, m, \{w_1, w_2\})-$*separating hash family*(or SHF) is a set of N functions Ψ, such that $|Y| = n, |X| = m, f : Y \to X$ for each $f \in \Psi$, and for any disjoint subsets $C_1, C_2 \subset Y$ with $|C_1| = w_1$ and $|C_2| = w_2$, there exists at least one $f \in \Psi$ such that $I_i := \{f(y) \mid y \in C_i\}, i \in \overline{1, 2}$ are also disjoint with each other. The following proposition shows that SHF is an equivalent concept with separating code in some sense.

Proposition 1.3 *An $(n, M)_q-$ code is $(w_1, w_2)-$separating if and only if its matrix representation is equal to the one of $(n; M, q, \{w_1, w_2\})-$SHF where $M \geq w_1 + w_2$.*

The bounds on SHF depend on values of certain chromatic polynomials. Given a graph $G = (V, E)$, the *chromatic polynomial* $\pi(G, m)$ is defined as follows : For a positive integer m, $\pi(G, m)$ denotes the number of $m-$colorings of G(i.e., the number of ways to color the vertices of G using a specified set of m colors, such that no two vertices $v_1, v_2 \in V$ having the same color are joined by an edge $e \in E$). It is well-known that $\pi(G, m)$ is a polynomial in m of degree $|V|$. Theorem 1.1 shows the properties of $\pi(G, m)$.

Theorem 1.1 (see [10]) *Let G be a graph on v vertices and ϵ edges. Then,*

- *the degree of $\pi(G, m)$ is v,*
- *the coefficient of m^v in $\pi(G, m)$ is 1,*
- *the coefficient of m^{v-1} in $\pi(G, m)$ is $-\epsilon$, and*
- *the coefficients of $\pi(G, m)$ alternate in sign.*

Reed-Solomon code, we will define below, is one of the most famous error-correcting codes and has an application in digital fingerprinting also.

Definition 1.2 Let α be a primitive element of $\mathbb{F}_q$. *Reed-Solomon code* $\mathrm{RS}_k(q)$ is a $[q-1, k]_q-$linear code $\mathrm{RS}_k(q) := \{ev(f) \mid f \in \mathbb{F}_q[x], \deg f \leq k-1\}$ where ev is an evaluation map $ev : f \mapsto (f(\alpha), f(\alpha^2), \cdots, f(\alpha^{q-1}))$.

In coding theory, it is well-known that Reed-Solomon code is Maximal Distance Seperable code that attains the Singleton Bound with equality. Thus, $d = q - k$ where d is minimum distance of $\mathrm{RS}_k(q)$.

1.3 Previous Works

Since separation property plays an important role in collusion-secure digital fingerprinting, lots of earlier works were done in the literature.

J. Staddon, et al. [12] studied the upper bound on the size of $w-$SFP codes $\Gamma \leq q^{\lceil n/w \rceil}+2w-2$. D. Stinson, et al. [14] improved the bound in case $w = 2$. Their result was $\Gamma \leq 4q^{\lceil n/3 \rceil} - 3$. D. Stinson, et al. [15] extended this bound to more general cases. The study was in SHF version and it claimed $\Gamma \leq (2w^2-3w+2)q^{\lceil n/(2w-1) \rceil} - 2w^2+3w-1$ for $w-$ SFP codes and $\Gamma \leq (2w^2-5w+4)q^{\lceil n/(2w-2) \rceil} - 2w^2+5w-3$ for $(w, w-1)-$ separating codes. The nonconstructive bounds, or existence conditions for separating codes have been studied. D. Stinson, et al. [13] achieved the following result.

Theorem 1.2 *Suppose n, m, w_1 and w_2 are all positive integers and $g = 1-\pi(\mathbb{K}^{w_1,w_2}, m)/m^{w_1+w_2}$ where $\mathbb{K}^{w_1,w_2}$ is the complete bipartite graph with w_1 vertices in one part and w_2 vertices in other part. If $N > w\log_g n^{-1}$ or $N > (w-1)\log_g(2n)^{-1}$, then there exists an $(N; n, m, \{w_1, w_2\})-$SHF.*

The similar study was done by D. Deng, et al. [4]. Their result is as follows.

Theorem 1.3 *There exists an $(N; n, m, \{w_1, w_2\})-$SHF if*

$$N > \frac{\log\left(e\left(\binom{n}{w_1}\binom{n-w_1}{w_2} - \binom{n-w_1-w_2}{w_1}\binom{n-2w_1-w_2}{w_2}\right)\right)}{-\log g}$$

where $g = 1 - \pi(\mathbb{K}^{w_1,w_2}, m)/m^{w_1+w_2}$.

Since restricted separation property provides colluders with more ability to attack, it has wider application than separation. But not as for separating codes, the bound on the size of restricted separating codes has not been widely studied. The only result on that topic we could find in the literature is Theorem 1.4 below by A. Barg, et al. [1].

Theorem 1.4 *Let $\alpha > 0$. Infinite sequences of restricted $(2, 2)-$separating codes exist for all rates R such that $R + \alpha \leq -3^{-1}\log_q(1 - (q-1)/q^3)$.*

For Reed-Solomon codes have certain collusion-secure properties, A. Silverberg, et al. [11] researched on applying Reed-Solomon codes and their list decoding algorithm to digital fingerprinting issue. During that discussion, they could gain some clues which led them to the following open problem.

Question 1 *Is it the case that $d > n - n/w^2$ for all $w-$IPP Reed-Solomon codes of length n and minimum distance d?*

By Proposition 1.1, we can see that if Question 1 is answered "yes", $w-$IPP is an equivalent condition with $w-$TA for Reed-Solomon code family. Question 1 was studied in [6], [9]. J. Moreira, et al. [9] proved that if w divides q, then $d > n - n/w^2$ for all $w-$SFP Reed-Solomon codes, therefore, $w-$SFP implies $w-$TA for Reed-Solomon

codes. M. Fernandez, et al. [6] worked on the same topic to show that if $n-d$ divides n, $(w_1, w_2)-$separation implies $d > n - n/(w_1 w_2)$. In that spirit, we can naturally suggest the question about necessary and sufficient condition of separation for Reed-Solomon codes as follows. (Sufficient condition comes from Proposition 1.2.)

Question 2 *Is it the case that $d > n - n/(w_1 w_2)$ for all $(w_1, w_2)-$separating Reed-Solomon codes of length n and minimum distance d?*

We begin in Section 2 by introducing an existence result for $(w_1, w_2)-$separating codes and discuss the "optimality" of the upper bounds. In Section 3, an upper bound on the size of restricted $(w, w)-$separating codes is obtained. We discuss Question 1 and Question 2 in Section 4. We will give some results about the conditions under which Question 2 holds positive answer both in general and intuitive versions. In the final Section, we will give conclusion and aspects for further works.

Through the rest of the paper, w_1 and w_2 are positive integers. $\mathbb{F}_q$ is a Galois field of characteristic p and order $q = p^m$. For an additive group G, the set $G\backslash\{0\}$ is denoted by G^*. For arbitrary two subsets $A, B \subset \mathbb{F}_q$, we define $AB := \{ab \mid a \in A, b \in B\}$ and $A + B := \{a + b \mid a \in A, b \in B\}$.

2 Existence Result for Separation

As we discussed in the previous section, the recent strong upper bounds on the size of $(w_1, w_2)-$separating codes have the identical form $|\Gamma| \leq f \cdot q^e + h$, where e, f and h are functions in w_1 and w_2. It's easy to see that the major contribution to the bound is due to q^e, when q increases, and h could be tolerated in some sense. D. Stinson, et al. [15] suggested the upper bound with $e = \lceil n/(w_1 + w_2 - 1) \rceil$. In this section we are attempting to answer the question *"How much is it possible to improve e?"*. The main result is Theorem 2.1 stated below.

Theorem 2.1 *For all real number $r < n/(w_1 + w_2 - 1)$, there exist infinite sequence of $(w_1, w_2)-$ separating $q-$ary codes of length n with more than q^r codewords.*

Proof. Set

$$\begin{aligned}
\psi(M) &:= \binom{M}{w_1}\binom{M-w_1}{w_2} - \binom{M-w_1-w_2}{w_1}\binom{M-2w_1-w_2}{w_2} \\
&= (w_1!w_2!)^{-1}(M(M-1)\cdots(M-w_1-w_2+1) - \\
&\quad (M-w_1-w_2)(M-w_1-w_2-1)\cdots(M-2w_1-2w_2+1))
\end{aligned}$$

If we expand $\psi(M)$, we get a polynomial of degree $w_1 + w_2 - 1$. The leading coefficient is $(w_1 + w_2)^2/v$ and the next coefficient is negative where $v = w_1!w_2!$. Therefore there exists M_0 such that $aM^{w-1} > (w_1 + w_2)^2 M^{w-1}/v > \psi(M)$ for all $M \geq M_0$ where $a = (w_1 + w_2)^2/(w_1 w_2)$.

Set $\rho(q) := q^{w_1+w_2} - \pi(\mathbb{K}^{w_1,w_2}, q)$. Theorem 1.1 implies that ρ is a polynomial in q

of degree w_1+w_2-1, its leading coefficient is the number of edges of $\mathbb{K}^{w_1,w_2}$, thus, w_1w_2 and the next coefficient is negative. Therefore, there exists q_0 such that $w_1w_2q^{w_1+w_2-1} > \rho(q)$ for all $q \geq q_0$.

Now set

$$\begin{aligned} \epsilon &:= n/(w_1+w_2-1) - r, \\ c &:= (e(w_1+w_2)^2(w_1w_2)^{n-1})^{1/(w_1+w_2-1)} \text{ and} \\ q' &:= \max\{q_0, [M_0^{1/r}]+1, [c^{1/\epsilon}]+1\}. \end{aligned}$$

Then, for all q and M such that $q \geq q', q^r < M < c^{-1}q^{n/(w_1+w_2-1)}$, there exist $(n, M)_q-$code which is $(w_1, w_2)-$separating. In fact, $q \geq q' > c^{1/\epsilon}$ implies $q^r + 1 < c^{-1}q^{n/(w_1+w_2-1)}$, so it is always possible to choose M as above. Moreover, $aM^{w-1} > \psi(M)$ since $M > q^r \geq (q')^r > M_0$, $w_1w_2q^{w_1+w_2-1} > \rho(q)$ for $q \geq q' \geq q_0$, and therefore the followings hold true.

$$\begin{aligned} e\psi(M) &< eaM^{w_1+w_2-1} \\ &< e \cdot (w_1+w_2)^2/(w_1w_2) \cdot q^n/(e \cdot (w_1+w_2)^2 \cdot (w_1w_2)^{n-1}) \\ &= (q/(w_1w_2))^n \\ &< (q^{w_1+w_2}/\rho(q))^n \end{aligned}$$

It meets the conditions of Theorem 1.3, thus, we get the conclusion. ⊡

From Theorem 2.1 we can see that e cannot be improved better than $n/(w_1+w_2-1)$, and $\lceil n/(w_1+w_2-1)\rceil$ obtained in [15] is "optimal" when we consider e integral. So reducing the amount of f seems promising in later studies for upper bounds.

3 Necessary Condition for Restricted Separation

In this section we propose a bound on the size of restricted $(w, w)-$separating codes. Note that the bound is independent on alphabet size q.

Theorem 3.1 *Let $w \geq 3$ be a positive integer. If C is a code of length n with M codewords and satisfies restricted $(w, w)-$separation property, then $M \leq 2^{\lfloor (n-w+2)/2 \rfloor} + w - 2$.*

Proof. Choose an arbitrary subset U of C with $w-2$ codewords, say $U = \{x^{(1)}, x^{(2)}, \cdots, x^{(w-2)}\}$. We can assume that the distance between $x^{(1)}$ and $x^{(2)}$ is larger than one without generality. (It can be easily checked that there exist such two codewords by using definition of restricted separation property.) Then, there exist two coordinates $i_0, i_1 \in \overline{1,n}$ such that $x^{(1)}_{i_0} \neq x^{(2)}_{i_0}$ and $x^{(1)}_{i_1} \neq x^{(2)}_{i_1}$. According to the definition of restricted separation property, $x^{(3)} \in \text{fea}(\{x^{(1)}, x^{(2)}\})$. Therefore, there exists a coordinate $i_2 \in \overline{1,n}$ such that $x^{(1)}_{i_2} = x^{(2)}_{i_2} \neq x^{(3)}_{i_2}$, and it is obvious that i_0, i_1, i_2 are different with each other. One can continue such steps to get different coordinates $i_0, i_1, \cdots, i_{w-3}$ where

$x_{i_k}^{(1)} = x_{i_k}^{(2)} \cdots = x_{i_k}^{(k)} \neq x_{i_k}^{(k+1)}$ for all $0 \leq k \leq w-3$. Thus, if we denote the number of all coordinates on which every codeword of U agrees by d, then $d \leq n-w+2$.

We can assume that the all the elements of U coincide on and only on the first d coordinates. Let us set $S := \{1, 2, \cdots, d\}$ and define a mapping $\Gamma(y) := \{i \in S \mid y_i = x_i^{(1)}\} \subset S$ for all $y \in C\backslash U$. If $y, z, t \in C\backslash U$ are distinct elements, then the followings are true : (1) $\Gamma(y) \cap \Gamma(z) \neq \phi$, (2) $\Gamma(y) \not\subset \Gamma(z)$, (3) $\Gamma(y) \cup \Gamma(z) \neq S$, (4) $\Gamma(y) \cap \Gamma(z) \not\subset \Gamma(t)$ and (5) $\Gamma(t) \not\subset \Gamma(y) \cap \Gamma(z)$ since their negations imply (a) fea$(U \cup \{y, z\}) = \mathbb{F}_q^n$, (b) fea$(U \cup \{y\}) \cap$fea$(U \cup \{z\}) = \{z\}$, (c) fea$(U) \cap$fea$(\{y, z\}) \neq \phi$, (d) fea$(U \cup \{y, z\}) \cap$fea$(\{t\}) = \{t\}$ and (e) fea$(U \cup \{t\}) \cap$fea$(\{y, z\}) \neq \phi$, respectively, which all contradict the restricted $(w, w)-$separation property.

CASE I : Assume that there exists $y^{(0)} \in C\backslash U$ such that $|\Gamma(y^{(0)})| \leq \lfloor d/2 \rfloor$. For all $y \in C\backslash U$, define the correspondence $\Gamma'(y) := \Gamma(y) \cap \Gamma(y^{(0)})$. Then Γ' is an injection since (4). For Γ' maps $C\backslash U$ to $\Gamma(y^{(0)})$ of at most $\lfloor d/2 \rfloor$ elements, we get $|C\backslash U| \leq 2^{\lfloor d/2 \rfloor}$.

CASE II : Assume that for all $y \in C\backslash U$, $|\Gamma(y)| > \lfloor d/2 \rfloor$. Set $\Gamma_1(y) := S\backslash\Gamma(y)$, then Γ_1 also satisfies (1) - (5). Similarly as above, we get $|C\backslash U| \leq 2^{\lfloor d/2 \rfloor}$.

From the two results above, $|C| = |U| + |C\backslash U| \leq 2^{\lfloor (n-w+2)/2 \rfloor} + w - 2$. $\square$

4 Separation for Reed-Solomon codes

In this section, we study Silverberg's open problem Question 1 and its generalized version Question 2. First, we state the following sufficient condition of non-separation for linear codes which can integrate the previous results in [6], [9].

Theorem 4.1 *Let C be an $[n, k]_q-$linear code containing* $\mathbf{1}= (1, 1, \cdots, 1)$. *Suppose there exists a codeword $c = (c_1, c_2, \cdots, c_n) \in C$ such that $V(c) \geq 2$ where $V(c) := \{c_i\}$ and there are subsets $E, F \subset V(c) (1 \leq |E| \leq w_1, 1 \leq |F| \leq w_2$ satisfying $V(c) = EF$ or $V(c) = E + F$. Then, C is not $(w_1, w_2)-$separating.*

Proof. We will only prove when $V(c) = E + F$, since the other case can be proven in similar way. Define $U := \{\beta \cdot \mathbf{1} \mid \beta \in E\}$ and $V := \{c - \gamma \cdot \mathbf{1} \mid \gamma \in F\}$. Then $U, V \subset C$ since $c, \mathbf{1} \in C$ and C is a linear code. Further, U and V are disjoint because $|V(c)| \geq 2$. For all $i \in \overline{1, n}$, there exist $\beta_i \in E$ and $\gamma_i \in F$ such that $c_i = \beta_i + \gamma_i$. If we set $x := (\beta_1, \beta_2, \cdots, \beta_n)$, then one can easily check that $x \in$ desc$U \cap$ desc$V$ which implies non-separation. $\square$

For Reed-Solomon code is a linear code containing **1**, we get the following corollary.

Corollary 4.1 *Let f be a non-constant polynomial of degree less than k and denote the image of f by* Imf. *Suppose there exist two subsets $E, F \subset$ Imf such that $1 \leq |E| \leq w_1, 1 \leq |F| \leq w_2$, and either* Im$f = E + F$ *or* Im$f = EF$ *is true. Then, the code* RS$_k(q)$ *is not $(w_1, w_2)-$separating.*

As one can see in the following examples, Theorem 4.1 or Corollary 4.1 includes the results in [6], [9] as special cases.

Example 1 : Suppose $n-d$ divides n, in other words, $k-1 \mid q-1$ for $\mathrm{RS}_k(q)$ where d is minimum distance. If the condition $d > n(1-1/(w_1w_2))$ fails, then $k-1 \geq (q-1)/(w_1w_2)$. Set $f := x^{k-1}$. Then f is a multiplicative homomorphism from $\mathbb{F}_q^*$ to $\mathbb{F}_q^*$. So $\mathrm{Im}f$ is a subgroup of $\mathbb{F}_q^*$ with order $|\mathrm{Im}f| = |\mathbb{F}_q^*|/|\mathrm{Ker}f| = (q-1)/(k-1) \leq w_1w_2$. For $\mathbb{F}_q^*$ is cyclic, $\mathrm{Im}f$ is also cyclic, thus, it has a generator γ. Set $E := \{\gamma^{iw_2} \mid 0 \leq i \leq w_1 - 1\}$ and $F := \{\gamma^j \mid 0 \leq j \leq w_2 - 1\}$. f, E, and F satisfy the condition of Corollary 4.1 so that $\mathrm{RS}_k(q)$ is not $(w_1, w_2)-$separating.

Example 2 : Assume $w \mid q$ and $k-1 \geq (q-1)/w^2$. The polynomial $f := x^{q/w^2} - x$ is an additive homomorphism over $\mathbb{F}_q$ and $|\mathrm{Im}f| = w^2$. By finite group theory, there exist subgroups $E, F < \mathrm{Im}f$ with w_1, w_2 elements such that $\mathrm{Im}f = E + F$. Thus, from Corollary 4.1 the code $\mathrm{RS}_k(q)$ is not $(w_1, w_2)-$separating.

The following theorem states a new parameter configuration for holding Question 2 positive.

Theorem 4.2 *Suppose $k-1 \mid q$ and at least one of the following conditions (1)-(3) are true. Then, $d > n(1-1/(w_1w_2))$ for all $(w_1, w_2)-$separating $\mathrm{RS}_k(q)$.*

(1) $k-1 \geq pq/(w_1w_2)$
(2) $(w_1/p^{r_1}) \cdot (w_2/p^{r_2}) < p$
(3) $[w_1/p^{r_1}] \cdot [w_2/p^{r_2}] \geq p$

where $r_i = [\log_p w_i]$ and d is minimum distance.

Proof. Let us suppose that $d \leq n(1-1/(w_1w_2))$, namely $k-1 \geq (q-1)/(w_1w_2)$ in spite of $(w_1, w_2)-$separation of $\mathrm{RS}_k(q)$ where $k-1 \mid q$. Set $f := x^{k-1} - x$. Then the similar discussion with Example 2 leads us to $|\mathrm{Im}f| = q/(k-1)$.

Assume (1) is true. Then $|\mathrm{Im}f| = q/(k-1) \leq w_1w_2/p \leq p^{r_1+r_2}$ and there exist $t_1, t_2 (t_1 \leq w_1, t_2 \leq w_2)$ such that $|\mathrm{Im}f| = p^{t_1+t_2}$ for $|\mathrm{Im}f|$ is a power of p. Using group theory, we find that there exist subgroups of $\mathrm{Im}f$, say E and F such that $\mathrm{Im}f = E + F$ and $|E| = p^{t_1}, |F| = p^{t_2}$. Thus, it contradicts the separation property from Corollary 4.1.

Assume (2) is true. We get $|\mathrm{Im}f| = q/(k-1) \leq w_1w_2 < p^{r_1+r_2+1}$ and since $|\mathrm{Im}f|$ is a power of p, $|\mathrm{Im}f| \leq p^{r_1+r_2}$. Then the exactly same argument as above holds.

Assume (1) and (2) are false, and (3) is true. Failure of (1) tells $q/(w_1w_2) \leq k-1 \leq pq/(w_1w_2)$ and $w_1w_2 > p^{r_1+r_2+1}$ since (2) is false and p is a prime number. Therefore, we could get $p^{r_1+r_2} < w_1w_2/p < |\mathrm{Im}f| = q/(k-1) < w_1w_2 < p^{r_1+r_2+2}$. So $|\mathrm{Im}f| = p^{r_1+r_2+1}$ for $|\mathrm{Im}f|$ is a power of p. Then there exist E, F and P with orders p^{r_1}, p^{r_2} and p, respectively, such that $\mathrm{Im}f = E+F+P$. Moreover, P is cyclic for its order is a prime number. We denote the generator of P by γ. Similarly with Example 1, set $P_1 := \{(i \cdot [w_2/p^{r_2}])\gamma \mid 0 \leq i \leq [w_1/p^{r_1}] - 1\}$ and $P_2 := \{(j\gamma \mid 0 \leq j \leq [w_2/p^{r_2}] - 1\}$. Since (3), we get $P = P_1 + P_2$ and $\mathrm{Im}f = (E + P_1) + (F + P_2)$. The cardinality of $E + P_1$ and $F + P_2$ are obviously not larger than w_1 and w_2, respectively. Therefore, by Corollary 4.1 we get the contradiction also in this case. $\square$

Proposition 4.1 *Suppose that* $w_1w_2 \geq q$ *or* $w_1w_2 \mid q-1$. *Then* $d > n(1-1/(w_1w_2))$ *for all* $(w_1, w_2)-$*separating* $\mathrm{RS}_k(q)$.

Proof. Suppose $w_1w_2 \geq q$. Then $(q-1)/(w_1w_2) < 1$. By Example 1, $\mathrm{RS}_2(q)$ is not $(w_1, w_2)-$separating. Since $\mathrm{RS}_k(q) \subset \mathrm{RS}_{k+1}(q)$ for all k, non-separation of $\mathrm{RS}_k(q)$ implies non-separation of $\mathrm{RS}_{k+1}(q)$. Therefore, if $k-1 \geq (q-1)/(w_1w_2)$, then $\mathrm{RS}_k(q)$ is not $(w_1, w_2)-$separating. In case when $w_1w_2 \mid q-1$ we can prove with the same way. $\square$

Experimental results allow us to answer Question 2 positive for a large family of Reed-Solomon codes. (J. Moreira, et al. [9] constructed a table illustrating some families of Reed-Solomon codes covered by positive answer using the results of Example 1 and Example 2.)

The statement of Theorem 4.1 and the proof of Theorem 4.2 give us a hint to describe the condition of "yes" on Question 2 in algebraic way. To show that, we will revisit the concept of subspace subcodes motivated and studied by P. Delsarte [3], J. Jensen [8], M. Hattori, et al. [7] and M. Dijk, et al [5]. Given an $[n,k]_q-$linear code C and a $v-$dimensional subspace $S(0 \leq v \leq m)$ of $\mathbb{F}_q$ where $q = p^m$, *subspace subcode* C_S of C is defined to be the set of codewords from C whose components all lie in S. It is obvious that C_S is a linear code over F_p. From the definition, the code spanned by **1** is a subspace subcode of dimension one. We call this code a *trivial subspace subcode*.

Before relating Question 2 to subspace subcodes, we define the following number theoretical function H_p where p is a prime number.

$$H_p(n) := \begin{cases} 0 & n < p^2 \text{ and } n \neq p \\ \log_p n & n \text{ is a power of } p \\ \lfloor \log_p n \rfloor - 1 & \text{otherwise} \end{cases}, \quad n \in \mathbb{N}$$

Theorem 4.3 *Suppose* $w_1w_2 < q = p^m$ *and* $v = H_p(w_1w_2) > 0$. *If there exist a* $v-$*dimensional subsace* S *of* $\mathbb{F}_q$ *and non-trivial subspace subcode* C_S *of* $C = \mathrm{RS}_k(q)$, *i.e.,* $\dim(C_S) \geq 2$, *then the parent code* $\mathrm{RS}_k(q)$ *is not* $(w_1, w_2)-$*separating.*

Proof. For C_S is not trivial, we can choose a codeword c in C_S which is not a multiple of **1**. Thus, $V(c) \subset S$, and $|V(c)| \geq 2$. Let us use the notations of Theorem 4.2. If w_1w_2 is a power of p, $|S| = p^v = w_1w_2$ and it is not difficult to check that the similar argument in Example 2 can be applied to show that $\mathrm{RS}_k(q)$ is not $(w_1, w_2)-$ separating. If w_1w_2 is not a power of p, then $p^v = p^{[\log_p n]-1} \leq p^{r_1+r_2}$. Therefore, we can prove non-separation of $\mathrm{RS}_k(q)$ in the similar way with the proof of Theorem 4.2 (1) or (2). $\square$

To make Theorem 4.3 valid, we must get the existence condition of non-trivial subspace subcode C_S which is linear. The dimension of subspace subcode has been studied in many works. M. Hattori, et al. [7] proposed an explicit formula as well as a simple lower bound for the dimension of subspace subcodes of Reed-Solomon codes where $p = 2$. For fixed m and v, they proved that most of subspace subcodes meet the very lower bound

with equality. They called the subspace on which dimensions of all the subcodes don't meet the lower bound *exceptional*. It is clear that exceptional subspaces are the ones we are looking for.

5 Conclusions

Properties of separating codes and the bounds on their cardinality, code length or dimension are important topics in digital fingerprinting code theory and so many problems remain open.

In this paper, we illustrated that in the upper bound formula on the size of $(w_1, w_2)-$separating codes the index of exponent should be larger than $n/(w_1 + w_2 - 1)$. Therefore, the improvement of the index seems to be accomplished between $n/(w_1 + w_2 - 1)$ and $\lceil n/(w_1 + w_2 - 1) \rceil$.

Restricted separation is quite strong condition so that the upper bound on the code size is supposed to be still smaller than the bound for separating codes. In that spirit, Theorem 3.1 leaves much to be desired.

Study on the relationship between separation and dimension for Reed-Solomon codes in the paper integrated the former results and some new conditions into an algebraic version using Theorem 4.1, and ended with the interesting result related to subspace subcodes. If we know the condition under which exceptional arises, Theorem 4.3 seems to be able to provide with a good hint for Silverberg's open problem.

Acknowledgement

We would like to thank J. Moreira for sending his article [9] by e-mail and offering advices on Silverberg's open problem.

References

[1] A. Barg and G. Kabatiansky, "Robust parent-identifying codes and combinatorial arrays", *online preprint*, 2011

[2] G. Cohen, "Separation and witnesses", *published in International Workshop on Coding and Cryptography (Hunan China)*, 2009

[3] P. Delsarte, "On subfield subcodes of modified Reed-Solomon codes", *IEEE Transactions on Information Theory*, **IT-21**(1975), 575-576

[4] D. Deng and D. Stinson, "The Lovász local lemma and its applications to some combinatorial arrays", *Designs, Codes and Cryptography*, **32**(2004), 121-134

[5] M. Dijk and L. Tolhuizen, "Efficient encoding for a class of subspace subcodes", *IEEE Transactions on Information Theory*, **45**(1999), 2142-2146

[6] M. Fernandez, J. Cotrina, M. Soriano and N. Domingo, ”A note about the identifier parent property in Reed-Solomon codes”, *Computers & Security*, **29**(2010), 628-635

[7] M. Hattori, R. McEliece and G. Solomon, ”Subspace subcodes of Reed-Solomon codes”, *IEEE Transactions on Information Theory*, **44**(1998), 1861-1880

[8] J. Jensen, ”Subgroup subcodes”, *IEEE Transactions on Information Theory*, **41**(1995), 781-785

[9] J. Moreira, M. Fernandez and M. Soriano, ”A note on the equivalence of the traceability properties of Reed-Solomon codes for certain coalition sizes”, *First IEEE Workshop on Information Forensics and Security (WIFS 2009)*, 36-40

[10] R. Reed, ”An introduction to chromatic polynomials”, *Journal of Combinatorial Theory*, **4**(1968), 52-71

[11] A. Silverberg, J. Staddon and J. Walker, ”Applications of list decoding to tracing traitors”, *IEEE Transactions on Information Theory*, **49**(2003), 1312-1318

[12] J. Staddon, D. Stinson and R. Wei, ”Combinatorial properties of frameproof and traceability codes”, *IEEE Transactions on Information Theory*, **47**(2001), 1042-1049

[13] D. Stinson, T. Trung and R. Wei, ”Secure frameproof codes, key distribution patterns, group testing algorithms and related structures”, *Journal of Statistical Planning and Inference*, **86**(2000), 595-617

[14] D. Stinson, R. Wei and K. Chen, ”On generalized separating hash families”, *Journal of Combinatorial Theory A*, **115**(2008), 105-120

[15] D. Stinson and G. Zaverucha, ”Some improved bounds for secure frameproof codes and related separating hash families”, *IEEE Transactions on Information Theory*, **54**(2008), 2508-2514